# Road Safety Rating & Optimal Route Evaluation

## Using Bicycle Accident Data


## Group members

**Zhou Yimin  Zhang Aiyin  Xu Ningxi  Gao He**



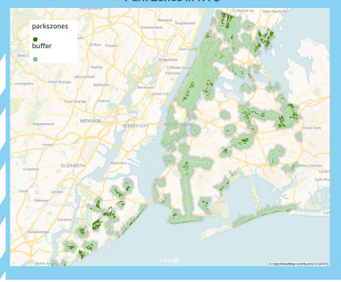
## Abstract

This project aims to provide a uniform safety rating for roads nearing parks across New York City by obtaining the significant features and their respective weights that contribute to crash accidents using ExtraTree regression model. Based on the results, an optimal safe route was evaluated for cyclists to a specific park in NYC.


## Introduction

The risk of cycling to a park may be higher due to the converging of more pedestrians and traffic volume in areas nearing parks. In recent years, many researchers have been studied the safety rating of roads, many countries or organizations have also developed road assessment programmes that mainly focus on monitoring the performance of various elements in the road networks and identifying deficiencies in road network design.

*Park Zones in NYC*




The main objectives of our project:

- To raise concerns about improving road safety;
- To provide a uniform safety rating for roads in areas nearing parks;
- To reduce deaths and injuries on roads by systematically assessing risk and identifying collision hotspots around parks;
- To propose safety assessment as the core of strategic decisions for route improvement, collision protection and route management standards.

## Workflow


## Data

| | |
|---|---|
| NYPD Motor Vehicle Collisions | NYC Speed Humps |
| NYC Street Centerline | NYC Speed Limits Signs |
| NYC Parks Zones | NYC Land Use |
| NYC Traffic Volume Counts (2014-2018) | |

## Method

- **Data Cleaning and Matching**

*-Network analysis:* The bicycle road network within 20 minutes from the park was screened out.

*-Buffer analysis:* Road attribute data within the study areas were cut .

*-Spatial join analysis:* New York City Bicycle Accident Data was exported and match to the road network by using spatial join tools. Using Pandas in Python to transform and match the road attribute data to the road network.

- **Pattern Analysis**

-Identifying Top 10 Contributors in Bicycle Collisions.

-Comparing Hot Spot Map.

-Discovering locations with most repeated bicycle collisions.

- **Regression Modeling**

-The significant features and their contribution to collisions were discovered.

-The collision counts per meter were calculated for each road segment, and outliers are drop out based on 1σ.

-Divide the count into three level: low, medium, high. The correlation coefficient matrix between each feature and collision count is generated.

-The ExtraTree Classifier was used for training model and predicting collision count level.

-The correct rate of prediction is obtained by comparing with the true value.

- **Road Safety Rating & Optimal Safe Route Evaluation**

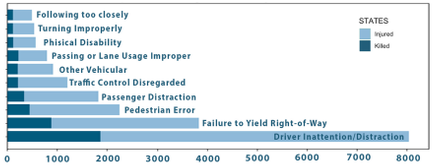
-The safety level is distributed to each segment of roads based on the prediction result of regression modelling.

-Determine optimal routes based on the network analysis tool in ArcGIS. The result is then compared to the general shortest path.

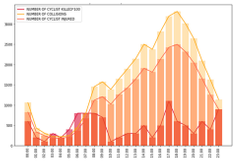
## Results

### Pattern analysis

*Bar plot of top 10 contributors in bicycle collisions*



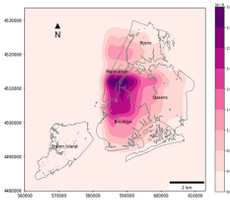

*Bicycle Collisions by 1 hour time interval*



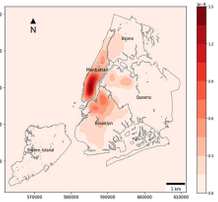
-The driver inattention/distraction was the main factor that leading to bicycle collisions.

-A couple noticeable trends that could warrant further investigation. First, there was gradual increasing trend of number of collisions until it reached a peak at 6 pm. Another noteworthy thing is that although the number of accidents is not high from 6 am. to 7 am, the fatality rate of the accident is very astonishing during this period.

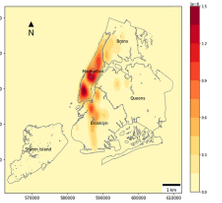
*Hot Spot Map of Collisions in NYC*



*Hot Spot Map of Bicycle Collisions in NYC*



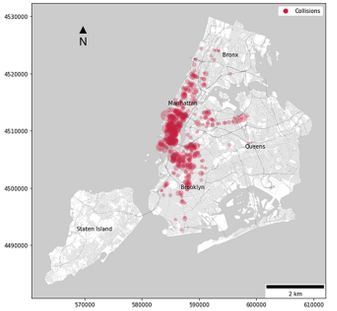
*Hot Spot Map of Bicycle Collisions Near Parks in NYC*



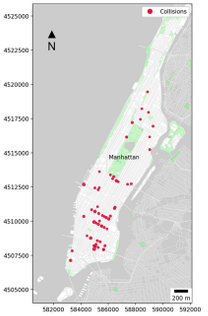
-Traffic accidents are concentrated in Manhattan Borough, the north part of Brooklyn Borough, and the west part of Queens Borough.

-The Washington Square Park, the Central Park, the Brooklyn Bridge Park, and the McCarren Park are high-risk parks areas of bicycle accidents.

*Top 500 Locations with Most Repeated bicycles Collisions*

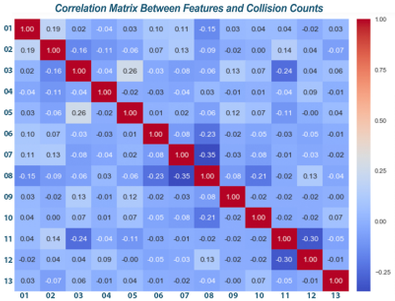



*Locations with Repeated Bicycle Collision Greater than 10*

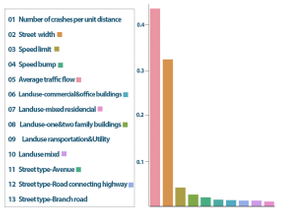



### Modeling Results

*Correlation Matrix Between Features and Collision Counts*



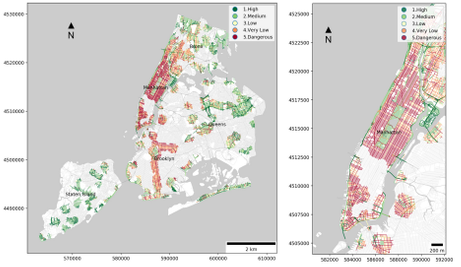
*Bar Plot of Feature Importance*




From the correlation coefficient matrix, the street width has a positive correlation with the number of crashes, and the correlation coefficient is 0.19. Followed by two land use types—mixed residential and commercial buildings with correlation coefficient equal to 0.11 and 0.10 respectively. The traffic volume per meter was the most significant feature in modelling result. Street width was the second important one followed by the street speed limit.

### Road Safety Rating

*Road Safety Levels within Park 20 minutes Buffers*

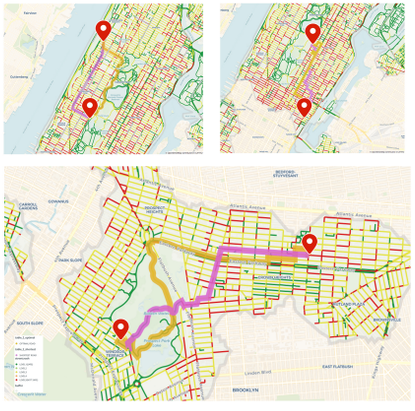



The maps that showing five safety levels of roads (high safety, medium safety, low safety, very low safety, dangerous). It can be seen that road risk in Manhattan is relatively high.

### Optimal Route to Park

The yellow path was the optimal safe route and the pink route was the general shortest route. By comparison, the optimal path we provide is safer and does not detour compared to the shortest path.

*The Optimal Route to Park*



## Conclusion

Some policy recommendations were made for the government:

1. To increase attention and increase enforcement of traffic rules for driver inattention/distraction as the main reason that leading to bicycle collisions.
2. Paying extra attention and strengthen traffic penalties to collisions during 6am. to 7am. as the fatality rate is extremely high.
3. For parks with frequent traffic crashes (such as Washington Square Park), strengthen education on civilians' safety awareness and increase road safety level nearby.
4. More relevant features about collisions require further exploration.


References

Ferreira, S., & Couto, A. (2012). Categorical Modeling to Evaluate Road Safety at the Planning Level. Journal of Transportation Safety & Security, 4(4), 308-322.

Jerrett, Su, Macleod, Hanning, Houston, & Wolch. (2016). Safe Routes to Play? Pedestrian and Bicyclist Crashes Near Parks in Los Angeles. Environmental Research, 151, 742-755.

The European Road Assessment Programme (EuroRAP)


Data Sources:

- (https://data.cityofnewyork.us/Public-Safety/NYPD-Motor-Vehicle-Collisions/h9gi-nx95)
- (https://data.cityofnewyork.us/City-Government/NYC-Street-Centerline-CSCL-/exjm-f27b)
- (https://data.cityofnewyork.us/City-Government/Parks-Zones/rjaj-zgq7)
- (https://data.cityofnewyork.us/Transportation/Traffic-Volume-Counts-2014-2018-/ertz-hr4r)
- (http://www.nycvzv.info/)

Know more about us